(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 033 405 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
06.09.2000 Bulletin 2000/36

(21) Application number: **00301439.6**

(22) Date of filing: **25.02.2000**

(51) Int Cl.[7]: **C12N 15/29**, C12N 15/82, C07K 14/415, C12Q 1/68, A01H 5/00

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: **25.02.1999 US 121825 P**
27.07.1999 US 145918 P
28.07.1999 US 145951 P
02.08.1999 US 146388 P
02.08.1999 US 146389 P
02.08.1999 US 146386 P
03.08.1999 US 147038 P
04.08.1999 US 147302 P
04.08.1999 US 147204 P
More priorities on the following pages

(83) **Declaration under Rule 28(4) EPC (expert solution)**

(71) Applicant: **Ceres Incorporated**
Malibu, CA 90265 (US)

(72) Inventors:
- **Alexandrov, Nickolai**
  Thousand Oaks, CA 91320 (US)
- **Brover, Vyacheslav**
  Calabasas, CA 91302 (US)
- **Chen, Xianfeng**
  Los Angeles, CA 90025 (US)
- **Subramanian, Gopalakrishnan**
  Moorpark, CA 93021 (US)
- **Troukhan, Maxim E.**
  South Pasadena, CA 91030 (US)
- **Zheng, Liansheng**
  Creve Coeur, MO 63141 (US)
- **Dumas, J.**
  , (US)

(74) Representative:
Bannerman, David Gardner et al
Withers & Rogers,
Goldings House,
2 Hays Lane
London SE1 2HW (GB)

Remarks:
THE COMPLETE DOCUMENT INCLUDING REFERENCE TABLES AND THE SEQUENCE LISTING IS AVAILABLE ON CD-ROM FROM THE EUROPEAN PATENT OFFICE, VIENNA SUB-OFFICE.

(54) Sequence-determined DNA fragments and corresponding polypeptides encoded thereby

(57) The present invention provides DNA molecules that constitute fragments of the genome of a plant, and polypeptides encoded thereby. The DNA molecules are useful for specifying a gene product in cells, either as a promoter or as a protein coding sequence or as an UTR or as a 3' termination sequence, and are also useful in controlling the behavior of a gene in the chromosome, in controlling the expression of a gene or as tools for genetic mapping, recognizing or isolating identical or related DNA fragments, or identification of a particular individual organism, or for clustering of a group of organisms with a common trait.

[0]Arabidopsis DNA is used in the present experiment, but the procedure is a general one.

Printed by Jouve, 75001 PARIS (FR)

(Cont. next page)

(30) Priority: Continued from first page
05.08.1999 US 147260 P
05.08.1999 US 147192 P
06.08.1999 US 147303 P
06.08.1999 US 147416 P
09.08.1999 US 147493 P
09.08.1999 US 147935 P
10.08.1999 US 148171 P
11.08.1999 US 148319 P
12.08.1999 US 148341 P
13.08.1999 US 148565 P
13.08.1999 US 148684 P
16.08.1999 US 149368 P
17.08.1999 US 149175 P
18.08.1999 US 149426 P
20.08.1999 US 149722 P
20.08.1999 US 149929 P
20.08.1999 US 149723 P
23.08.1999 US 149902 P
23.08.1999 US 149930 P
25.08.1999 US 150566 P
26.08.1999 US 150884 P
27.08.1999 US 151065 P
27.08.1999 US 151066 P
27.08.1999 US 151080 P
30.08.1999 US 151303 P
31.08.1999 US 151438 P
01.09.1999 US 151930 P
07.09.1999 US 152363 P
10.09.1999 US 153070 P
13.09.1999 US 153758 P
15.09.1999 US 154018 P
16.09.1999 US 154039 P
20.09.1999 US 154779 P
22.09.1999 US 155139 P
23.09.1999 US 155486 P
24.09.1999 US 155659 P
28.09.1999 US 156458 P
29.09.1999 US 156596 P
04.10.1999 US 157117 P
05.10.1999 US 157753 P
06.10.1999 US 157865 P
07.10.1999 US 158029 P
08.10.1999 US 158232 P
12.10.1999 US 158369 P
13.10.1999 US 159294 P
13.10.1999 US 159295 P
13.10.1999 US 159293 P
14.10.1999 US 159638 P
14.10.1999 US 159637 P
14.10.1999 US 159329 P
14.10.1999 US 159331 P
14.10.1999 US 159330 P
18.10.1999 US 159584 P
21.10.1999 US 160815 P
21.10.1999 US 160767 P
21.10.1999 US 160768 P
21.10.1999 US 160741 P
21.10.1999 US 160770 P
21.10.1999 US 160814 P
22.10.1999 US 160981 P
22.10.1999 US 160980 P
22.10.1999 US 160989 P
25.10.1999 US 161405 P
25.10.1999 US 161404 P
25.10.1999 US 161406 P
26.10.1999 US 161361 P
26.10.1999 US 161360 P
26.10.1999 US 161359 P
28.10.1999 US 161920 P
28.10.1999 US 161992 P
28.10.1999 US 161993 P
29.10.1999 US 162143 P
29.10.1999 US 162142 P
29.10.1999 US 162228 P
01.11.1999 US 162895 P
01.11.1999 US 162891 P
01.11.1999 US 162894 P
02.11.1999 US 163093 P
02.11.1999 US 163092 P
02.11.1999 US 163091 P
03.11.1999 US 163249 P
03.11.1999 US 163248 P
03.11.1999 US 163281 P
04.11.1999 US 163380 P
04.11.1999 US 163381 P
04.11.1999 US 163379 P
08.11.1999 US 164151 P
08.11.1999 US 164150 P
08.11.1999 US 164146 P
09.11.1999 US 164260 P
09.11.1999 US 164259 P
10.11.1999 US 164548 P
10.11.1999 US 164317 P
10.11.1999 US 164321 P
10.11.1999 US 164318 P
10.11.1999 US 164544 P
10.11.1999 US 164545 P
10.11.1999 US 164319 P
12.11.1999 US 164870 P
12.11.1999 US 164959 P
12.11.1999 US 164962 P
12.11.1999 US 164960 P
12.11.1999 US 164871 P
12.11.1999 US 164961 P
15.11.1999 US 164927 P
15.11.1999 US 164929 P
15.11.1999 US 164926 P
16.11.1999 US 165669 P
16.11.1999 US 165671 P
16.11.1999 US 165661 P
17.11.1999 US 165919 P
17.11.1999 US 165918 P
17.11.1999 US 165911 P
18.11.1999 US 166158 P
18.11.1999 US 166157 P

(Cont. next page)

18.11.1999 US 166173 P
19.11.1999 US 166412 P
19.11.1999 US 166419 P
19.11.1999 US 166411 P
22.11.1999 US 166733 P
22.11.1999 US 166750 P
23.11.1999 US 167362 P
24.11.1999 US 167382 P
24.11.1999 US 167233 P
24.11.1999 US 167234 P
24.11.1999 US 167235 P
30.11.1999 US 167904 P
30.11.1999 US 167908 P
30.11.1999 US 167902 P
01.12.1999 US 168232 P
01.12.1999 US 168233 P
01.12.1999 US 168231 P
02.12.1999 US 168546 P
02.12.1999 US 168549 P
02.12.1999 US 168548 P
03.12.1999 US 168673 P
03.12.1999 US 168675 P
03.12.1999 US 168674 P
07.12.1999 US 169278 P
07.12.1999 US 169302 P
07.12.1999 US 169298 P
08.12.1999 US 169692 P
08.12.1999 US 169691 P
16.12.1999 US 171107 P
16.12.1999 US 171098 P
16.12.1999 US 171114 P
19.01.2000 US 176866 P
19.01.2000 US 176867 P
19.01.2000 US 176910 P
26.01.2000 US 178166 P
27.01.2000 US 178547 P
27.01.2000 US 177666 P
27.01.2000 US 178546 P
27.01.2000 US 178544 P
27.01.2000 US 178545 P
28.01.2000 US 178755 P
28.01.2000 US 178754 P
01.02.2000 US 179395 P
01.02.2000 US 179388 P
03.02.2000 US 180039 P
03.02.2000 US 180139 P
04.02.2000 US 180207 P
04.02.2000 US 180206 P
07.02.2000 US 180695 P
07.02.2000 US 180696 P
09.02.2000 US 181228 P
09.02.2000 US 181214 P
10.02.2000 US 181476 P
10.02.2000 US 181551 P
15.02.2000 US 182477 P
15.02.2000 US 182516 P
15.02.2000 US 182512 P
15.02.2000 US 182478 P

17.02.2000 US 183165 P
17.02.2000 US 183166 P
27.07.1999 US 145913 P
05.03.1999 US 123180 P
09.03.1999 US 123548 P
23.03.1999 US 125788 P
25.03.1999 US 126264 P
29.03.1999 US 126785 P
01.04.1999 US 127462 P
06.04.1999 US 128234 P
08.04.1999 US 128714 P
16.04.1999 US 129845 P
19.04.1999 US 130077 P
21.04.1999 US 130449 P
23.04.1999 US 130891 P
23.04.1999 US 130510 P
28.04.1999 US 131449 P
30.04.1999 US 132407 P
30.04.1999 US 132048 P
04.05.1999 US 132484 P
05.05.1999 US 132485 P
06.05.1999 US 132487 P
06.05.1999 US 132486 P
07.05.1999 US 132863 P
11.05.1999 US 134256 P
14.05.1999 US 134221 P
14.05.1999 US 134218 P
14.05.1999 US 134370 P
14.05.1999 US 134219 P
18.05.1999 US 134768 P
19.05.1999 US 134941 P
20.05.1999 US 135124 P
21.05.1999 US 135353 P
24.05.1999 US 135629 P
25.05.1999 US 136021 P
27.05.1999 US 136392 P
28.05.1999 US 136782 P
01.06.1999 US 137222 P
03.06.1999 US 137528 P
04.06.1999 US 137502 P
07.06.1999 US 137724 P
08.06.1999 US 138094 P
10.06.1999 US 138540 P
10.06.1999 US 138847 P
14.06.1999 US 139119 P
16.06.1999 US 139452 P
16.06.1999 US 139453 P
17.06.1999 US 139492 P
18.06.1999 US 139461 P
18.06.1999 US 139750 P
18.06.1999 US 139463 P
18.06.1999 US 139457 P
18.06.1999 US 139459 P
18.06.1999 US 139462 P
18.06.1999 US 139455 P
18.06.1999 US 139458 P
18.06.1999 US 139454 P
18.06.1999 US 139456 P

(Cont. next page)

18.06.1999 US 139460 P
18.06.1999 US 139763 P
21.06.1999 US 139817 P
22.06.1999 US 139899 P
23.06.1999 US 140354 P
23.06.1999 US 140353 P
24.06.1999 US 140695 P
28.06.1999 US 140823 P
29.06.1999 US 140991 P
30.06.1999 US 141287 P
01.07.1999 US 142154 P
01.07.1999 US 141842 P
02.07.1999 US 142055 P
06.07.1999 US 142390 P
08.07.1999 US 142803 P
09.07.1999 US 142920 P
12.07.1999 US 142977 P
13.07.1999 US 143542 P
14.07.1999 US 143624 P
15.07.1999 US 144005 P
16.07.1999 US 144085 P
16.07.1999 US 144086 P
19.07.1999 US 144333 P
19.07.1999 US 144335 P
19.07.1999 US 144325 P
19.07.1999 US 144334 P
19.07.1999 US 144332 P
19.07.1999 US 144331 P
20.07.1999 US 144884 P
20.07.1999 US 144352 P
20.07.1999 US 144632 P
21.07.1999 US 144814 P
21.07.1999 US 145086 P
21.07.1999 US 145088 P
22.07.1999 US 145192 P
22.07.1999 US 145085 P
22.07.1999 US 145089 P
22.07.1999 US 145087 P
23.07.1999 US 145145 P
23.07.1999 US 145224 P
23.07.1999 US 145218 P
26.07.1999 US 145276 P
27.07.1999 US 145919 P

[1] Neuwald A.F., York J.D., Majerus P.W. FEBS Lett. 294:16-18(1991).
[2] Glaeser H-U., Thomas D., Gaxiola R., Montrichard F., Surdin-Kerjan Y., Serrano R. EMBO J. 12:3105-3110 (1993).
[3] Bone R., Springer J.P., Atack J.R. Proc. Natl. Acad. Sci. U.S.A. 89:10031-10035(1992).

[0924] 313. Ion transport protein

[0925] This family contains Sodium, Potassium, Calcium ion channel This family is 6 transmembrane helices in which the last two helices flank a loop which determines ion selectivity. In some sub-families (e.g. Na channels) the domain is repeated four times, whereas in others (e.g. K channels) the protein forms as a tetramer in the membrane. A bacterial structure of the protein is known for the last two helices but is not the Pfam family due to it lacking the first four helices

[0926] 314. Isocitrate and isopropylmalate dehydrogenases signature (isodh)

Isocitrate dehydrogenase (IDH) [1,2] is an important enzyme of carbohydrate metabolism which catalyzes the oxidative decarboxylation of isocitrate into alpha-ketoglutarate. IDH is either dependent on NAD+ (EC 1.1.1.41) or on NADP+ (EC 1.1.1.42). In eukaryotes there are at least three isozymes of IDH: two are located in the mitochondrial matrix (one NAD+-dependent, the other NADP+-dependent), while the third one (also NADP+-dependent) is cytoplasmic. In Escherichia coli the activity of a NADP+-dependent form of the enzyme is controlled by the phosphorylation of a serine residue; the phosphorylated form of IDH is completely inactivated. 3-isopropylmalate dehydrogenase (EC 1.1.1.85) (IMDH) [3,4] catalyzes the third step in the biosynthesis of leucine in bacteria and fungi, the oxidative decarboxylation of 3-isopropylmalate into 2-oxo-4-methylvalerate. Tartrate dehydrogenase (EC 1.1.1.93) [5] catalyzes the reduction of tartrate to oxaloglycolate. These enzymes are evolutionary related [1,3,4,5]. The best conserved region of these enzymes is a glycine-rich stretch of residues located in the C-terminal section. This region was used as a signature pattern.

[0927] Consensus pattern: [NS]-[LIMYT]-[FYDN]-G-[DNT]-[IMVY]-x-[STGDN]-[DN]-x(2)-[SGAP]-x(3,4)-G-[STG]-[LIVMPA]-G-[LIVMF]-

[1] Hurley J.H., Thorsness P.E., Ramalingam V., Helmers N.H., Koshland D.E. Jr., Stroud R.M. Proc. Natl. Acad. Sci. U.S.A. 86:8635-8639(1989).
[2] Cupp J.R., McAlister-Henn L. J. Biol. Chem. 266:22199-22205(1991).
[3] Imada K., Sato M., Tanaka N., Katsube Y., Matsuura Y., Oshima T. J. Mol. Biol. 222:725-738(1991).
[4] Zhang T., Koshland D.E. Jr. Protein Sci. 4:84-92(1995).
[5] Tipton P.A., Beecher B.S. Arch. Biochem. Biophys. 313:15-21(1994).

[0928] 315. Jacalin-like lectin domain.

[0929] Proteins containing this domain are lectins. It is found in 1 to 6 copies in these proteins. The domain is also found in the animal prostatic spermine-binding protein (Swiss:P15501).

[0930] [1] Sankaranarayanan R, Sekar K, Banerjee R, Sharma V, Surolia A, Vijayan M; Nat Struct Biol 1996;3: 596-603.

[0931] 316. KH domain

[0932] KH motifs probably bind RNA directly. Auto antibodies to Nova, a KH domain protein, cause paraneoplastic opsoclonus ataxia.

[1] Burd CG, Dreyfuss G, Science 1994;265:615-621.
[2] Musco G, Stier G, Joseph C, Castiglione Morelli MA, Nilges M, Gibson TJ, Pastore A, Cell 1996;85:237-245.

[0933] 317. Kelch motif

[0934] The kelch motif was initially discovered in Kelch (Swiss:Q04652). In this protein there are six copies of the motif. It has been shown that Swiss:Q04652 is related to Galactose Oxidase [1] for which a structure has been solved [2]. The kelch motif forms a beta sheet. Several of these sheets associate to form a beta propeller structure as found in neur,

[0935] [1] Bork P, Doolittle RF, J Mol Biol 1994;236:1277-1282. [2] Ito N, Phillips SE, Stevens C, Ogel ZB, McPherson MJ, Keen, JN, Yadav KD, Knowles PF, Nature 1991;350:87-90.

[0936] 318. Soybean trypsin inhibitor (Kunitz) protease inhibitors family signature

[0937] The soybean trypsin inhibitor (Kunitz) family [1] is one of the numerous families of proteinase inhibitors. It comprise plant proteins which have inhibitory activity against serine proteinases from the trypsin and subtilisin families, thiol proteinases and aspartic proteinases as well as some proteins that are probably involved in seed storage. This family is currently known to group the following proteins: - Trypsin inhibitors A, B, C, KTI1, and KTI2 from soybean. - Trypsin inhibitor DE3 from coral beans (Erythrina sp.). - Trypsin inhibitor DE5 from sandal bead tree. - Trypsin inhibitors 1A (WTI-1A), 1B (WTI-1B), and 2 (WTI-2) from goa bean. - Trypsin inhibitor from Acacia confusa. - Trypsin inhibitor from silk tree. - Chymotrypsin inhibitor 3 (WCI-3) from goa bean. - Cathepsin D inhibitors PDI and NDI from potato [2],

which inhibit both cathepsin D (aspartic proteinase) and trypsin. - Alpha-amylase/subtilisin inhibitors from barley and wheat. - Albumin-1 (WBA-1) from goa bean seeds [3]. - Miraculin from Richadella dulcifica [4], a sweet taste protein. - Sporamin from sweet potato [5], the major tuberous root protein. - Thiol proteinase inhibitor PCPI 8.3 (P340) from potato tuber [6]. - Wound responsive protein gwin3 from poplar tree [7]. - 21 Kd seed protein from cocoa [8]. All these proteins contain from 170 to 200 amino acid residues and one or twointrachain disulfide bonds. The best conserved region is found in their N-terminal section and is used as a signature pattern

[0938] Consensus pattern: [LIVM]-x-D-x-[EDNTY]-[DG]-[RKHDENQ]-x-[LIVM]-x(5)-Y-x-[LIVM] -

[1] Laskowski M., Kato I. Annu. Rev. Biochem. 49:593-626(1980).
[2] Ritonja A., Krizaj I., Mesko P., Kopitar M., Lucovnik P., Strukelj B., Pungercar J., Buttle D.J., Barrett A.J., Turk V. FEBS Lett. 267:13-15(1990).
[3] Kortt A.A., Strike P.M., de Jersey J. Eur. J. Biochem. 181:403-408(1989).
[4] Theerasilp S., Hitotsuya H., Nakajo S., Nakaja K., Nakamura Y., Kurihara Y. J. Biol. Chem. 264:6655-6659 (1989).
[5] Hattori T., Yoshida N., Nakamura K. Plant Mol. Biol. 13:563-572(1989).
[6] Krizaj I., Drobnic-Kosorok M., Brzin J., Jerala R., Turk V. FEBS Lett. 333:15-20(1993).
[7] Bradshaw H.D., Hollick J.B., Parsons T.J., Clarke H.R.G., Gordon M.P. Plant Mol. Biol. 14:51-59(1989).
[8] Tai H., McHenry L., Fritz P.J., Furtek D.B. Plant Mol. Biol. 16:913-915(1991).

[0939] 319. Beta-ketoacyl synthases active site

Beta-ketoacyl-ACP synthase (KAS) [1] is the enzyme that catalyzes the condensation of malonyl-ACP with the growing fatty acid chain. It is found as a component of the following enzymatic systems: - Fatty acid synthetase (FAS), which catalyzes the formation of long-chain fatty acids from acetyl-CoA, malonyl-CoA and NADPH. Bacterial and plant chloroplast FAS are composed of eight separate subunits which correspond to different enzymatic activities; beta-ketoacyl synthase is one of these polypeptides. Fungal FAS consists of two multifunctional proteins, FAS1 and FAS2; the beta-ketoacyl synthase domain is located in the C-terminal section of FAS2. Vertebrate FAS consists of a single multifunctional chain; the beta-ketoacyl synthase domain is located in the N-terminal section [2]. - The multifunctional 6-methysalicylic acid synthase (MSAS) from Penicillium patulum [3]. This is a multifunctional enzyme involved in the biosynthesis of a polyketide antibiotic and which has a KAS domain in its N-terminal section. - Polyketide antibiotic synthase enzyme systems. Polyketides are secondary metabolites produced by microorganisms and plants from simple fatty acids. KAS is one of the components involved in the biosynthesis of the Streptomyces polyketide antibiotics granatacin [4], tetracenomycin C [5] and erythromycin. - Emericella nidulans multifunctional protein Wa. Wa is involved in the biosynthesis of conidial green pigment. Wa is protein of 216 Kd that contains a KAS domain. - Rhizobium nodulation protein nodE, which probably acts as a beta-ketoacyl synthase in the synthesis of the nodulation Nod factor fatty acyl chain. - Yeast mitochondrial protein CEM1. The condensation reaction is a two step process: the acyl component of an activated acyl primer is transferred to a cysteine residue of the enzyme and is then condensed with an activated malonyl donor with the concomitant release of carbon dioxide. The sequence around the active site cysteine is well conserved and can be used as a signature pattern.

[0940] Consensus pattern: G-x(4)-[LIVMFAP]-x(2)-[AGC]-C-[STA](2)-[STAG]-x(3)-[LIVMF] [C is the active site residue]

[1] Kauppinen S., Siggaard-Andersen M., von Wettstein-Knowles P. Carlsberg Res. Commun. 53:357-370(1988).
[2] Witkowski A., Rangan V.S., Randhawa Z.I., Amy C.M., Smith S. Eur. J. Biochem. 198:571-579(1991).
[3] Beck J., Ripka S., Siegner A., Schiltz E., Schweizer E. Eur. J. Biochem. 192:487-498(1990).
[4] Bibb M.J., Biro S., Motamedi H., Collins J.F., Hutchinson C.R. EMBO J. 8:2727-2736(1989).
[5] Sherman D.H., Malpartida F., Bibb M.J., Kieser H.M., Bibb M.J., Hopwood D.A. EMBO J. 8:2717-2725(1989).

[0941] 320. Kinesin motor domain signature and profile

Kinesin [1,2,3] is a microtubule-associated force-producing protein that mayplay a role in organelle transport. Kinesin is an oligomeric complex composedof two heavy chains and two light chains. The kinesin motor activity isdirected toward the microtubule's plus end.The heavy chain is composed of three structural domains: a large globular N-terminal domain which is responsible for the motor activity of kinesin (it isknown to hydrolyze ATP, to bind and move on microtubules), a central alpha-helical coiled coil domain that mediates the heavy chain dimerization; and asmall globular C-terminal domain which interacts with other proteins (such asthe kinesin light chains), vesicles and membranous organelles.A number of proteins have been recently found that contain a domain similarto that of the kinesin 'motor' domain [1,4,E1]: - Drosophila claret segregational protein (ncd). Ncd is required for normal chromosomal segregation in meiosis, in females, and in early mitotic divisions of the embryo. The ncd motor activity is directed toward the microtubule's minus end. - Drosophila kinesin-like protein (nod). Nod is required for the distributive chromosome segre-

1/1 - (C) FILE REGISTRY
RN - 301764-80-7 REGISTRY
CN - Protein (Arabidopsis thaliana clone Ceres_1715205) (9CI) (CA INDEX NAME
OTHER NAMES:
CN - 938: PN: EP1033405 SEQID: 34187 claimed protein
FS - PROTEIN SEQUENCE
SQL - 421

SEQ
1 LIVQWLREKR VKKHMASLPL GPQPHALAPP LQLHDGDALK RRPELDSDKE
51 MSAAVIEGND AVTGHIISTT IGGKNGEPKQ TISYMAERVV GTGSFGIVFQ
101 AKCLETGESV AIKKVLQDRR YKNRELQLMR PMDHPNVISL KHCFFSTTSR
151 DELFLNLVME YVPETLYRVL RHYTSSNQRM PIFYVKLYTY QIFRGLAYIH
201 TVPGVCHRDV KPQNLLVDPL THQVKLCDFG SAKVLVKGEP NISYICSRYY
251 RAPELIFGAT EYTASIDIWS AGCVLAELLL GQPLFPGENS VDQLVEIIKV
301 LGTPTREEIR CMNPNYTDFR FPQIKAHPWH KVFHKRMPPE AIDLASRLLQ
351 YSPSLRCTAL EACAHPFFNE LREPNARLPN GRPLPPLFNF KQELGGASME
401 LINRLIPEHV RRQMSTGLQN S

SEQ3
1 Leu-Ile-Val-Gln-Trp-Leu-Arg-Glu-Lys-Arg-
11 Val-Lys-Lys-His-Met-Ala-Ser-Leu-Pro-Leu-
21 Gly-Pro-Gln-Pro-His-Ala-Leu-Ala-Pro-Pro-
31 Leu-Gln-Leu-His-Asp-Gly-Asp-Ala-Leu-Lys-
41 Arg-Arg-Pro-Glu-Leu-Asp-Ser-Asp-Lys-Glu-
51 Met-Ser-Ala-Ala-Val-Ile-Glu-Gly-Asn-Asp-
61 Ala-Val-Thr-Gly-His-Ile-Ile-Ser-Thr-Thr-
71 Ile-Gly-Gly-Lys-Asn-Gly-Glu-Pro-Lys-Gln-
81 Thr-Ile-Ser-Tyr-Met-Ala-Glu-Arg-Val-Val-
91 Gly-Thr-Gly-Ser-Phe-Gly-Ile-Val-Phe-Gln-
101 Ala-Lys-Cys-Leu-Glu-Thr-Gly-Glu-Ser-Val-
111 Ala-Ile-Lys-Lys-Val-Leu-Gln-Asp-Arg-Arg-
121 Tyr-Lys-Asn-Arg-Glu-Leu-Gln-Leu-Met-Arg-
131 Pro-Met-Asp-His-Pro-Asn-Val-Ile-Ser-Leu-
141 Lys-His-Cys-Phe-Phe-Ser-Thr-Thr-Ser-Arg-
151 Asp-Glu-Leu-Phe-Leu-Asn-Leu-Val-Met-Glu-
161 Tyr-Val-Pro-Glu-Thr-Leu-Tyr-Arg-Val-Leu-
171 Arg-His-Tyr-Thr-Ser-Ser-Asn-Gln-Arg-Met-
181 Pro-Ile-Phe-Tyr-Val-Lys-Leu-Tyr-Thr-Tyr-
191 Gln-Ile-Phe-Arg-Gly-Leu-Ala-Tyr-Ile-His-
201 Thr-Val-Pro-Gly-Val-Cys-His-Arg-Asp-Val-
211 Lys-Pro-Gln-Asn-Leu-Leu-Val-Asp-Pro-Leu-
221 Thr-His-Gln-Val-Lys-Leu-Cys-Asp-Phe-Gly-
231 Ser-Ala-Lys-Val-Leu-Val-Lys-Gly-Glu-Pro-
241 Asn-Ile-Ser-Tyr-Ile-Cys-Ser-Arg-Tyr-Tyr-
251 Arg-Ala-Pro-Glu-Leu-Ile-Phe-Gly-Ala-Thr-
261 Glu-Tyr-Thr-Ala-Ser-Ile-Asp-Ile-Trp-Ser-
271 Ala-Gly-Cys-Val-Leu-Ala-Glu-Leu-Leu-Leu-
281 Gly-Gln-Pro-Leu-Phe-Pro-Gly-Glu-Asn-Ser-
291 Val-Asp-Gln-Leu-Val-Glu-Ile-Ile-Lys-Val-
301 Leu-Gly-Thr-Pro-Thr-Arg-Glu-Glu-Ile-Arg-
311 Cys-Met-Asn-Pro-Asn-Tyr-Thr-Asp-Phe-Arg-
321 Phe-Pro-Gln-Ile-Lys-Ala-His-Pro-Trp-His-
331 Lys-Val-Phe-His-Lys-Arg-Met-Pro-Pro-Glu-

341 Ala-Ile-Asp-Leu-Ala-Ser-Arg-Leu-Leu-Gln-
351 Tyr-Ser-Pro-Ser-Leu-Arg-Cys-Thr-Ala-Leu-
361 Glu-Ala-Cys-Ala-His-Pro-Phe-Phe-Asn-Glu-
371 Leu-Arg-Glu-Pro-Asn-Ala-Arg-Leu-Pro-Asn-
381 Gly-Arg-Pro-Leu-Pro-Pro-Leu-Phe-Asn-Phe-
391 Lys-Gln-Glu-Leu-Gly-Gly-Ala-Ser-Met-Glu-
401 Leu-Ile-Asn-Arg-Leu-Ile-Pro-Glu-His-Val-
411 Arg-Arg-Gln-Met-Ser-Thr-Gly-Leu-Gln-Asn-
421 Ser

Unspecified
MAN
CA
STN Files: CA, CAPLUS
1 REFERENCES IN FILE CA (1967 TO DATE)
1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

FERENCE 1

133:318291 CA
Sequence-determined DNA fragments and corresponding encoded polypeptides from corn and Arabidopsis
Alexandrov, Nickolai; Brover, Vyacheslav; Chen, Xianfeng; Subramanian, Gopalakrishnan; Troukhan, Maxim E.; Zheng, Liansheng; Dumas, J.
Ceres Inc., USA
Eur. Pat. Appl., 339 pp.
CODEN: EPXXDW
Patent
English
C12N015-29; C12N015-82; C07K014-415; C12Q001-68; A01H005-00
3-3 (Biochemical Genetics)
Section cross-reference(s): 6, 11
N.CNT 16

	PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
	EP 1033405	A2	20000906	EP 2000-301439	20000225
	R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO				
AI	US 1999-121825			19990225	
	US 1999-145918			19990727	
	US 1999-145951			19990728	
	US 1999-146386			19990802	
	US 1999-146388			19990802	
	US 1999-146389			19990802	
	US 1999-147038			19990803	
	US 1999-147204			19990804	
	US 1999-147302			19990804	
	US 1999-147192			19990805	
	US 1999-147260			19990805	
	US 1999-147303			19990806	
	US 1999-147416			19990806	
	US 1999-147493			19990809	
	US 1999-147935			19990809	
	US 1999-148171			19990810	
	US 1999-148319			19990811	
	US 1999-148341			19990812	
	US 1999-148565			19990813	
	US 1999-148684			19990813	

The present invention provides DNA mols. that constitute fragments of the genome and cDNAs from Zea mays mays (HYBRID SEED #35A19) and Arabidopsis

thaliana (ecotype Wassilewski), and polypeptides encoded thereby. The DNA mols. are useful for specifying a gene product in cells, either as a promoter or as a protein coding sequence or as an UTR or as a 3' termination sequence, and are also useful in controlling the behavior of a gene in the chromosome, in controlling the expression of a gene or as tools for genetic mapping, recognizing or isolating identical or related DNA fragments, or identification of a particular individual organism, or for clustering of a group of organisms with a common trait. Arabidopsis DNA is used in the present expt., but the procedure is a general one. Protocols are provided for Southern hybridizations and transformation of carrot cells. [This abstr. record is one of 15 records supplemental to CA13316218528Q necessitated by the large no. of index entries required to fully index the document and publication system constraints.].

corn Arabidopsis cDNA genome protein sequence